Homework 2

Jordan Rhoades

Due 2/13/2019

Instructions: Complete the following problems using RMarkdown to render a Microsoft Word document. Include code in the space provided. If a direct question is asked, answer that question below the code input for that problem.

## Problem 1

The following data represent the pulse rates (beats per minute) of nine students enrolled in a section Introductory Statistics course. Treat the nine students as a population.

|  |  |
| --- | --- |
| Student | Pulse |
| Perpectual Bempah | 76 |
| Megan Brooks | 60 |
| Jeff Honeycutt | 60 |
| Clarice Jefferson | 81 |
| Crystal Kurtenbach | 72 |
| Janette Lantka | 80 |
| Kevin McCarthy | 80 |
| Tammy Ohm | 68 |
| Kathy Wojdyla | 73 |

1. Determine the population mean pulse.

pulses <- c(76,60,60,81,72,80,80,68,73)  
mean(pulses)

## [1] 72.22222

1. Generate three simple random samples of size 3 and determine the sample mean pulse of each sample.

s1 <- sample(pulses,3)  
s2 <- sample(pulses,3)  
s3 <- sample(pulses,3)  
m1 <- mean(s1)  
m2 <- mean(s2)  
m3 <- mean(s3)  
m1

## [1] 69.33333

m2

## [1] 71.33333

m3

## [1] 71.33333

sprintf("Sample One's mean: %f, Sample Two's mean: %f, and Sample Three's mean: %f ",m1,m2,m3)

## [1] "Sample One's mean: 69.333333, Sample Two's mean: 71.333333, and Sample Three's mean: 71.333333 "

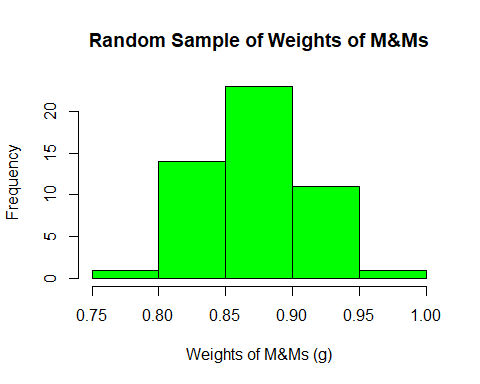
## Problem 2

The following data represent the weights (in grams) of a simple random sample of 50 M&M plain candies.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0.87 | 0.88 | 0.82 | 0.90 | 0.84 |
| 0.84 | 0.87 | 0.88 | 0.86 | 0.86 |
| 0.86 | 0.94 | 0.91 | 0.89 | 0.83 |
| 0.91 | 0.88 | 0.90 | 0.88 | 0.91 |
| 0.95 | 0.91 | 0.88 | 0.88 | 0.92 |
| 0.86 | 0.87 | 0.86 | 0.89 | 0.84 |
| 0.85 | 0.87 | 0.93 | 0.89 | 0.79 |
| 0.93 | 0.84 | 0.84 | 0.86 | 0.81 |
| 0.90 | 0.82 | 0.83 | 0.84 | 0.87 |
| 0.85 | 0.91 | 0.99 | 0.93 | 0.85 |

1. Determine the shape of the distribution of the weights of M&Ms by drawing a frequency histogram.

weight <- c( 0.87 ,0.88 ,0.82 ,0.90 ,0.84 ,  
 0.84 ,0.87 ,0.88 ,0.86 ,0.86 ,  
 0.86 ,0.94 ,0.91 ,0.89 ,0.83 ,  
 0.91 ,0.88 ,0.90 ,0.88 ,0.91 ,  
 0.95 ,0.91 ,0.88 ,0.88 ,0.92 ,  
 0.86 ,0.87 ,0.86 ,0.89 ,0.84 ,  
 0.85 ,0.87 ,0.93 ,0.89 ,0.79 ,  
 0.93 ,0.84 ,0.84 ,0.86 ,0.81 ,  
 0.90 ,0.82 ,0.83 ,0.84 ,0.87 ,  
 0.85 , 0.91 ,0.99 ,0.93 ,0.85)  
hist(weight,xlab="Weights of M&Ms (g)",main="Random Sample of Weights of M&Ms",col="green")



## The shape of the histogram is approximately symmetric so it's bell-shaped.

1. Find the mean and median.

sm <- mean(weight)  
med <- median(weight)  
sprintf("The mean weight of the random sample of M&Ms is %f grams and the median is %f grams",sm,med)

## [1] "The mean weight of the random sample of M&Ms is 0.875800 grams and the median is 0.870000 grams"

1. Which measure of central tendency better describes the weight of a plain M&M? ## Both because the median is typically the best indicator of central tendency especially when the data is skewed (not this time though). Since the data is symmetric, the mean works just as well since there aren’t outliers to through it off.

## Problem 3

Mr. Zuro finds the mean height of all 14 students in his class to be 68.0 inches. Just as Mr. Zuro finishes explaining how to get to the mean, Danielle walks in late. Danielle is 65 inches tall. What is the mean height of the 15 students in the class?

#We can multiply the first number of students with the mean, add the 15th student's height, then find the mean for all 15 students by having the new number of students divide the new summation.  
n1 <- 14  
mea1 <- 68.0  
sum1 <- n1 \* mea1  
sum2 <- sum1 + 65  
n2 <- 15  
mea2 <- sum2 / n2  
sprintf("The mean height of the 15 students is: %f inches.",mea2)

## [1] "The mean height of the 15 students is: 67.800000 inches."

## Problem 4

The following data represent the travel time (in minutes) to school for nine students. Treat the nine students as a population.

|  |  |
| --- | --- |
| Student | Travel Time |
| Amanda | 39 |
| Amber | 21 |
| Tim | 9 |
| Mike | 32 |
| Nicole | 30 |
| Scot | 45 |
| Erica | 11 |
| Tiffany | 12 |
| Glenn | 39 |

1. Determine the population standard deviation and the population variance.

tt <- c(39,21,9,32,30,45,11,12,39)  
N1 <- 9  
sig1 <- sd(tt) \* sqrt((N1 - 1)/N1)  
v1 <- var(tt) \* (N1 -1)/N1  
sprintf("The population standard deviation is %f and the population variance is %f.",sig1,v1)

## [1] "The population standard deviation is 12.841868 and the population variance is 164.913580."

1. Generate three simple random samples of size 4. Determine the sample standard deviation of each sample.

s4 <- sample(tt,4)  
s5 <- sample(tt,4)  
s6 <- sample(tt,4)  
sam1 <- sd(s4)  
sam2 <- sd(s5)  
sam3 <- sd(s6)  
sprintf("The sample standard deviations are Sample 1: %f, Sample 2: %f Sample 3: %f.",sam1,sam2,sam3)

## [1] "The sample standard deviations are Sample 1: 13.892444, Sample 2: 13.200379 Sample 3: 9.763879."

## Problem 5

The Stanford-Binet Intelligence Quotient (IQ) measures intelligence. IQ scores have a bell-shaped distribution with a mean of 100 and standard deviation of 15.

1. What percentage of people has an IQ score between 70 and 130?

#34+34+13.5+13.5=95%  
print("The percentage of people between 70 and 130 (2 sigma from the mean going -2 sigma to +2 sigma) is 95%.")

## [1] "The percentage of people between 70 and 130 (2 sigma from the mean going -2 sigma to +2 sigma) is 95%."

1. What percentage of people has an IQ score less than 70 or greater than 130?

print("The percentage of people with an IQ score less than 70 or greater than 130 is (100% - 95% =) 5%.")

## [1] "The percentage of people with an IQ score less than 70 or greater than 130 is (100% - 95% =) 5%."

1. What percentage of people has an IQ score greater than 85 and less than 145?

print("The percentage of people with an IQ score greater than 85 and less than 145 is (34% + 34% + 13.5% + 2.35% =) 83.85%")

## [1] "The percentage of people with an IQ score greater than 85 and less than 145 is (34% + 34% + 13.5% + 2.35% =) 83.85%"

## Problem 6

According to the U.S. Census Bureau, the mean commute time to work for a resident of Boston, Massachusetts, is 27.3 minutes. Assume that the standard deviation of the commute time is 8.1 minutes to answer the following:

1. What minimum percentage of commuters in Boston has a commute time within 2 standard deviations of the mean?

#34+34+13.5+13.5=95%  
print("The minimum percentage of commuters with a commute time witin 2 sigma is 95%.")

## [1] "The minimum percentage of commuters with a commute time witin 2 sigma is 95%."

1. What minimum percentage of commuters in Boston has a commute time within 1.5 standard deviations of the mean? What are the commute times within 1.5 standard deviations of the mean?

print("The minimum percent of commuters with a commute time within 1.5 sigma is 81.5% and the commute time within 1.5 sigma is within range of 19.2 minutes to 35.4 minutes.")

## [1] "The minimum percent of commuters with a commute time within 1.5 sigma is 81.5% and the commute time within 1.5 sigma is within range of 19.2 minutes to 35.4 minutes."

## Problem 7

Babies born after a gestation period of 32 - 35 weeks have a mean weight of 2600 grams and a standard deviation of 660 grams. Babies bord after a gestation period of 40 weeks have a mean weight of 3500 grams and a standard deviation of 470 grams. Suppose a 34-week gestation period baby weighs 3000 grams and a 40-week gestation period baby weight 3900 grams.

1. What is the -score for the 34-week gestation period baby?

#z-score = (x - mean)/std  
z2 <- (3000 - 2600)/660  
sprintf("The z-score for the 34-week gestation period is %f.",z2)

## [1] "The z-score for the 34-week gestation period is 0.606061."

1. What is the -score for the 40-week gestation period baby?

#z-score = (x - mean)/std  
z3 <- (3900 - 3500)/470  
sprintf("The z-score for the 40-week gestation period is %f.",z3)

## [1] "The z-score for the 40-week gestation period is 0.851064."

1. Which baby weighs less relative to the gestation period?

## The 40-week gestation period baby weighs more relative to the gestation period.

## Problem 8

A highly selective boarding school will only admit students who place at least 1.5 standard deviations above the mean on a standardized test that has a mean of 200 and a standard deviation of 26. What is the minimum score that an applicant must make on the test to be accepted?

std <- 26  
mean1 <- 200  
minsc <- mean1 + std \* 1.5  
sprintf("The minimum score that an applicant must make on the test to be accepted is %f.",minsc)

## [1] "The minimum score that an applicant must make on the test to be accepted is 239.000000."

## Problem 9

The data below preresent the miles per gallon of a random sample of SMART cars with a three-cylinder, 1.0 liter engine.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 31.5 | 36.0 | 37.8 | 38.4 | 40.1 | 42.3 |
| 34.3 | 36.3 | 37.9 | 38.8 | 40.6 | 42.7 |
| 34.5 | 37.4 | 38.0 | 39.3 | 41.4 | 43.5 |
| 35.5 | 37.5 | 38.3 | 39.5 | 41.4 | 47.5 |

1. Compute the -score corresponding to the individual who obtained 36.3 miles per gallon. Interpret this result.

mpg <- c(31.5, 36.0, 37.8, 38.4, 40.1, 42.3,   
34.3, 36.3, 37.9, 38.8, 40.6, 42.7,   
34.5, 37.4, 38.0, 39.3, 41.4, 43.5,   
35.5, 37.5, 38.3, 39.5, 41.4, 47.5)  
z1 <- (36.3 - mean(mpg))/sd(mpg)  
z1a <- abs(z1)  
sprintf("The z-score for this individual is %f which means it is %f standard deviations below the mean.",z1,z1a)

## [1] "The z-score for this individual is -0.723931 which means it is 0.723931 standard deviations below the mean."

1. Calculate and interpret the quartiles.

sort(mpg)

## [1] 31.5 34.3 34.5 35.5 36.0 36.3 37.4 37.5 37.8 37.9 38.0 38.3 38.4 38.8  
## [15] 39.3 39.5 40.1 40.6 41.4 41.4 42.3 42.7 43.5 47.5

quantile(mpg,type=2)

## 0% 25% 50% 75% 100%   
## 31.50 36.85 38.35 41.00 47.50

print("0% shows that 31.50 mpg is the minimum. 25% of the mpg data are less than or equal to 36.85 mpg (first quartile) and 75% of the data is greater than the 36.84 mpg. 50% shows that half of the data is less than or equal to 38.35 mpg (2nd quartile) and the other half of the data is greater than 38.35 mpg. 75% of the data is less than or equal to 41.00 mpg (third quartile) and 25% of the data are greater than 41.00 mpg. The 100% shows that the maximum value in the data set is 47.50")

## [1] "0% shows that 31.50 mpg is the minimum. 25% of the mpg data are less than or equal to 36.85 mpg (first quartile) and 75% of the data is greater than the 36.84 mpg. 50% shows that half of the data is less than or equal to 38.35 mpg (2nd quartile) and the other half of the data is greater than 38.35 mpg. 75% of the data is less than or equal to 41.00 mpg (third quartile) and 25% of the data are greater than 41.00 mpg. The 100% shows that the maximum value in the data set is 47.50"

1. Calculate and interpret the 10th percentile.

qua1 <- quantile(mpg, probs=.1)#output:=34.8  
sprintf("The tenth percentile is %f mpg and this means that 10 percent of the mpg data is less than or equal to %f mpg and that 90 percent of the data is greater than %f mpg.",qua1,qua1,qua1)

## [1] "The tenth percentile is 34.800000 mpg and this means that 10 percent of the mpg data is less than or equal to 34.800000 mpg and that 90 percent of the data is greater than 34.800000 mpg."

1. Compute the interquartile range.

#eqn:= IQR = Q\_3 - Q\_1  
q1 <- quantile(mpg, type = 2)[2]  
q3 <- quantile(mpg, type = 2)[4]  
q3-q1

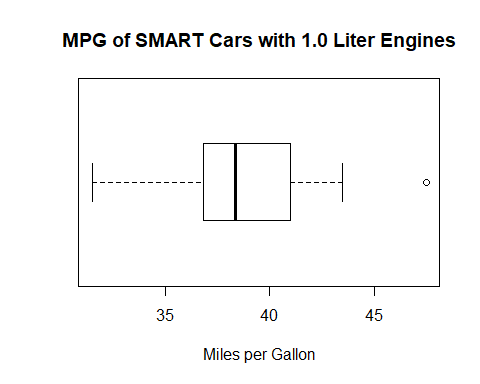
## 75%   
## 4.15

iqr <- IQR(mpg, type=2)  
sprintf("The interquartile range (range of the middle 50 percent of the data set) is %f.",iqr)

## [1] "The interquartile range (range of the middle 50 percent of the data set) is 4.150000."

1. Construct a boxplot of the data.

boxplot(mpg,horizontal=TRUE,main="MPG of SMART Cars with 1.0 Liter Engines",xlab="Miles per Gallon")



1. Determine the upper and lower fences. Are there any outliers present in this data set?

lf <- q1 - 1.5\*(iqr)  
uf <- q3 + 1.5\*(iqr)  
sum4 <- sum(mpg < lf | mpg > uf)  
sprintf("There is %i outlier in this data set.", sum4)

## [1] "There is 1 outlier in this data set."

## Problem 10

Mensa is an organization designed for people of high intelligence. One qualifies for Mensa if one’s intelligence is measured at or above the 98th percentile. Explain what this means.

## This means that someone who qualifies for this organization has at least an intelligence that is greater than or equal to 98% of those who have had their intelligence measured .